Amendments to the Specification:

Please replace paragraphs 0007 and 0008 with the following amended paragraph:

[0007] The manual interpretation of spectra, called *de novo* sequencing, is an approach that can sequence peptides without using database-searching programs ~~(Johnson, R. S. "How to sequence tryptic peptides using low energy CID data", http://www.abrf.org/ResearchGroups/MassSpectrometry/EPosters/ms97quiz/Sequencing Tutorial.html).~~ MS/MS spectra commonly contain short series of fragment ions where the mass differences between these ions match the masses of amino acids in the original peptide. These mass differences can be linked together to form partial or complete peptide sequences (McCormack, A. L.; Eng, J. K.; Yates, J. R. III *Methods Companion Methods Enzymol.* 1994, 6, 284-303). Areas of MS/MS spectra that cannot be assigned to standard amino acids may be due to incomplete peptide fragmentation, or to post-translational modifications that change the mass of amino acids. The manual interpretation of spectra is time consuming and requires considerable expertise. Fortunately, there are several commercial (Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* 2003, 17, 2337-2342; Scigelova, M.; Maroto, F.; Dufresne, C; Vazquez, J. "High-Throughput De Novo Sequencing", 14th Meeting Methods of Protein Structure Analysis, Valencia, Spain, September 8-12, 2002; Langridge, J. I.; Millar, A.; Young, P.; O'Malley, R.; Swainston, N.; Skilling, J.; Hoyes, J.; Richardson, K. "A Fully Automated Hierarchical Software Strategy for De Novo Sequencing of Whole Q-Tof Electrospray LC-MS/MS Datasets", Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, Florida, June 2-6, 2002) and freely available (Fernandez-de-Cossio, J.; Gonzalez, J.; Betancourt, L.; Besada, V.; Padron, G.; Shimonishi, Y.; Takao, T. *Rapid Commun. Mass Spectrom.* 1998, 12, 1867-1878; Taylor, J. A.; Johnson, R. S. *Anal. Chem.* 2001, 73, 2594-2604; Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, S.; Zerial, M.; Wilm, M. *Proteomics* 2001, 1, 668-682; Lu, B.; Chen, T. *J. Comp. Biol.* 2003, 10, 1-12) software packages that perform automated *de novo* sequencing. These

programs take into consideration much of the possible variation in peptide fragmentation, and introduce the possibility of high-throughput, objective MS/MS sequencing.

Please replace paragraph 0058 with the following amended paragraph:

[0058] Peptide matches with alignment scores over 85 are accepted as correct identifications. Example peptide matches with their corresponding alignments and alignment scores can be found ~~in a supplementary file on the web (Additional results and analysis can be found in the supplementary file on the web at http://medir.ohsu.edu/~geneview/ publication/opensea/)~~ on the World Wide Web. Peptides with long sequences typically have larger scores, however, due to the requirements placed on the actual generation of the alignments, long sequences are generally more difficult to match, justifying their higher score. We've found that factoring the peptide length into the scoring function does not significantly improve the separation of correct from incorrect matches.

Please replace paragraph 0067 with the following amended paragraph:

[0067] In this example, all MS/MS spectra acquired were *de novo* sequenced. Peaks 1.3 (Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* 2003, 17, 2337-234 ;Ma, B.; Zhang, K.; Liang. C. "An Effective Algorithm for the Peptide De Novo Sequencing from MS/MS Spectrum", The 14th Symposium on Combinatorial Pattern Matching, March 2003, 266-278) (Bioinformatics Solutions Inc., Waterloo, ON Canada) and Lutefisk1900 1.3.2 (Fernandez-de-Cossio, J.; Gonzalez, J.; Betancourt, L.; Besada, V.; Padron, G.; Shimonishi, Y.; Takao, T. *Rapid Commun. Mass Spectrom.* 1998, 12, 1867-1878; Current versions of Lutefisk ~~are available for download at http://www.hairyfatguy.com/Lutefisk/~~ can be found on the World Wide Web) *de novo* sequencing programs were used to test the performance of the methods and systems of the present invention. Both programs were configured to assume that all cysteines were alkylated and that all peptides were tryptically digested. Unlike Lutefisk, Peaks

reports full amino acid sequences without unknown mass regions, but does assign each amino acid in the sequence a confidence score. Sequence regions where amino acids had confidences scores below 50% were replaced by the combined mass of those amino acids. Lutefisk reports as many as five *de novo* sequences for each spectrum. All of these sequences were used to produce a match. Only the top scoring sequence reported by Peaks was used, as generally all of the top five Peaks sequences could be represented by the 50% consensus sequence.

Please replace paragraph 0073 with the following amended paragraph:

[0073] In one specific embodiment of the methods and systems of the present invention, the default alignment score cutoff of 85 identified 94% of the correct assignments (sensitivity) and eliminated 97% of the incorrect assignments (specificity). For comparison, the sensitivity of the Xcorr score used by SEQUEST was 77%, while the specificity was 85% using minimum Xcorr values of 1.8, 2.5, and 3.5 for peptides of +1, +2, and +3 charge, respectively (Figure 8b). Similarly, the sensitivity of the CIDentify E-value score was 70% and the specificity was 89% with a minimum score cutoff of 10^{-4} (Figure 8c). Statistical analysis of the methods and systems of the present invention alignment score distributions can be found in the supplementary file on the web ~~(Additional results and analysis can be found in the supplementary file on the web at http://medir.ohsu.edu/~geneview/publication/opensea/)~~.

Please incorporate the enclosed Sequence Listing in the specification on the page preceding the claims.

Please replace the abstract with the following amended abstract, which contains 136 words:

Methods for identifying a macromolecule having a sequence and sequence modifications thereof from mass spectrometry data, including a method comprising: providing at least one *de novo* sequence from mass spectroscopy data of sequences of fragments of said macromolecule, calculating at least one mass-based alignment between each *de novo* sequence and sequence in a sequence database, comparing

the *de novo* sequence of a fragment containing a modification site with a sequence in the database, identifying a difference by utilizing a modification catalog, calculating at least one match score for the mass-based alignment, identifying sequences in said database from mass-based alignment in response to the match score, and grouping identification of sequences from at least one *de novo* sequence into an identified macromolecule list that agrees with the mass, and storing the result on a computer readable medium.